

RESEARCH

Open Access



Globally elevated levels of histone H3 lysine 9 trimethylation in early infancy are associated with poor growth trajectory in Bangladeshi children

Kristyna Kupkova^{1,2}, Savera J. Shetty¹, Marilyn G. Pray-Grant³, Patrick A. Grant³, Rashidul Haque⁴, William A. Petri Jr.⁵ and David T. Auble^{1*}

Abstract

Background Stunting is a global health problem affecting hundreds of millions of children worldwide and contributing to 45% of deaths in children under the age of five. Current therapeutic interventions have limited efficacy. Understanding the epigenetic changes underlying stunting will elucidate molecular mechanisms and likely lead to new therapies.

Results We profiled the repressive mark histone H3 lysine 9 trimethylation (H3K9me3) genome-wide in peripheral blood mononuclear cells (PBMCs) from 18-week-old infants (n = 15) and mothers (n = 14) enrolled in the PROVIDE study established in an urban slum in Bangladesh. We associated H3K9me3 levels within individual loci as well as genome-wide with anthropometric measurements and other biomarkers of stunting and performed functional annotation of differentially affected regions. Despite the relatively small number of samples from this vulnerable population, we observed globally elevated H3K9me3 levels were associated with poor linear growth between birth and one year of age. A large proportion of the differentially methylated genes code for proteins targeting viral mRNA and highly significant regions were enriched in transposon elements with potential regulatory roles in immune system activation and cytokine production. Maternal data show a similar trend with child's anthropometry; however, these trends lack statistical significance to infer an intergenerational relationship.

Conclusions We speculate that high H3K9me3 levels may result in poor linear growth by repressing genes involved in immune system activation. Importantly, changes to H3K9me3 were detectable before the overt manifestation of stunting and therefore may be valuable as new biomarkers of stunting.

Keywords Stunting, Childhood undernutrition, Epigenetics, Histone methylation, Heterochromatin, Endogenous retroviruses

*Correspondence:

David T. Auble
auble@virginia.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Childhood stunting is a result of chronic undernutrition, and even with the available resources, stunting remains a global health problem affecting over 20% of children under the age of five worldwide [1]. Stunting is defined by a height-for-age z (HAZ) score below -2 and emerges within the first 1000 days post conception [1, 2]. Without treatment, stunting leads to a significant risk of mortality in children under the age of five and is associated with severe lifelong health and socioeconomic problems. Common issues associated with stunting include immune system dysfunction [3, 4], cognitive impairment [5, 6], poverty, and paradoxically excessive weight gain later in life often resulting in obesity and with it diseases such as type 2 diabetes and cardiovascular diseases [7, 8]. While nutrient limitation plays a significant role in the development of stunting, it is not the sole driver. Many factors contribute to stunting, including socioeconomic factors [9–12], limited access to clean water and poor sanitation, which lead to an increased pathogen exposure and can result in enteric infections with or without diarrhea altering gut health, triggering chronic inflammation, and negatively affecting microbiome maturation [13–17]. Maternal health has been identified as another key factor in the development of stunting [16, 18]. Mothers of stunted children were often stunted themselves and evidence suggests they can pass this condition to neonates before birth [19] through mechanisms that are not well understood but that can impact subsequent generations [20, 21].

One of the global nutrition targets released in 2014 is to achieve a 40% reduction in stunting in children under five by 2025 [22]. While there has been a decline in the rate of stunting since then, the current interventions are not effective enough to reach this goal. Multiple studies have conducted metabolomic or metagenomic profiling in association with child growth to identify key metabolites for new nutritional therapies [13, 23–26]. In our approach, we aim to understand the detailed changes to molecular circuitry underlying stunting to gain increased precision in targeting specific pathways as new therapies. We focus on epigenetic profiling, specifically targeting histone modifications. Epigenetic profiling provides a powerful tool for studying stunting, as the epigenome is responsive to and reflects environmental signals [27]. This is because the enzymes that deposit and remove histone marks are dependent on the availability of key metabolites, the levels of which are reflected in epigenetic profiles [28], and furthermore, it has been suggested that the epigenome plays a role in intragenerational and transgenerational inheritance [29, 30]. While multiple studies have been conducted to understand the epigenetic changes associated with malnutrition, most

of them have concentrated on DNA methylation [31–34] and histone modifications have been less well explored.

In our previous work, we showed that both H3 trimethylation on lysine 4 (H3K4me3, associated with active promoters [35]) and H3 acetylation on lysine 27 (H3K27ac, associated with active enhancers [36]) undergo large scale changes within the first year of life. We observed increased levels of activation at stress and immune response genes in 18-week-old children through hyperactivated H3K27ac. This was followed by massive changes in one-year-old stunted children in which H3K27ac was globally downregulated and H3K4me3 redistributed from promoters to distal ectopic sites. Both H3K4me3 and H3K27ac changes in one-year-olds point to downregulated immune responses and suggest alterations in one-carbon metabolism. In both cases we observed relatively modest changes in early infancy followed by large-scale changes at one year. H3K4me3 and H3K27ac are associated with active transcription and dynamic changes in response to environmental changes. This prompted us to explore the role of more stable chromatin structures in early infancy relative to the development of stunting. To address this, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) in 18-week-old children targeting H3 trimethylation on lysine 9 (H3K9me3). H3K9me3 mark is associated with constitutive heterochromatin and has a role in development by repressing lineage inappropriate genes, as well as a role in repressing endogenous retroviruses (ERVs) and other transposable elements (TEs) [37, 38], whose regulatory roles in gene expression are currently being investigated [39–41]. Furthermore, animal studies suggest that H3K9me3 plays a crucial role in transgenerational inheritance in response to environmental changes, including nutritional changes [42, 43]. For this reason, we also profiled H3K9me3 in mothers to uncover relationships between maternal and child heterochromatin in stunting. The results suggest that globally elevated H3K9me3 in early infancy predisposes children to poor growth, possibly through suppression of TEs involved in development of immune responses to pathogens. The H3K9me3 pattern in early infancy is thus a predictor of the risk of stunting and supports a model in which the immune system of stunted children becomes compromised in early infancy.

Results

Study overview

To identify genome-wide H3K9me3 changes associated with childhood stunting, we performed ChIP-seq targeting H3K9me3 in peripheral blood mononuclear cells (PBMCs) from 18-week-old infants (n = 15) from Bangladesh who were enrolled in the PROVIDE (performance

of rotavirus and oral polio vaccines in developing countries) study ($n=700$, ref. [44], Fig. 1a, Additional file 1: Table S1). While the blood samples come from 18-week-old infants, for each child we have anthropometric measurements including height-for-age z (HAZ) and weight-for-age z (WAZ) scores within the first year of the child's life with additional biomarkers from 18 weeks (Fig. 1b, Additional file 1: Table S2). Each HAZ score provides a snapshot of a child's current health status, which can change dynamically within the first year of life (Fig. 1c). In our analysis, we used the Δ HAZ score as a measure of growth change between two time points (Fig. 1d). Samples selected for this study reliably represent the PROVIDE spectrum of HAZ and Δ HAZ scores and are balanced in sex representation (Additional file 1: Fig. S1,2).

H3K9me3 levels are correlated with poor growth

The average H3K9me3 signal within identified H3K9me3 enriched regions, i.e., peaks, was increased in children with poor growth between birth and one year as measured by the Δ HAZ (birth:1yr) score (Fig. 2a). And the pattern was conserved between the two sexes (Additional

file 1: Fig. S3a,b). The Δ HAZ (birth:1yr) score showed the strongest correlation with the average H3K9me3 signal and H3K9me3 was negatively correlated with the Δ HAZ (birth:1yr) score (Fig. 2b). Furthermore, the first component (PC1) of principal component analysis (PCA) explained almost 80% of the variance in H3K9me3 peak signal (Fig. 2c) and again was strongly correlated with Δ HAZ (birth:1yr) score (Fig. 2d). While there were sex differences in the H3K9me3 profiles, these were associated with the second principal component (PC2) and explain only about 10% of the variance in the data. These results were not influenced by enzymatic cleavage ("clipping") of histone H3 [45], or by global changes to histone H3 levels (Additional file 1: Fig. S4, Additional file 1: Supplemental text).

Both the average H3K9me3 profiles and signal within H3K9me3 peaks used for PCA were normalized based on *Drosophila* spike-in content, which we used in our experimental design to uncover any global H3K9me3 profile changes (Additional file 1: Fig. S5a). While normalizing the H3K9me3 profiles based on read count per million mapped reads still showed a negative correlation between a child's Δ HAZ (birth:1yr) score and the overall

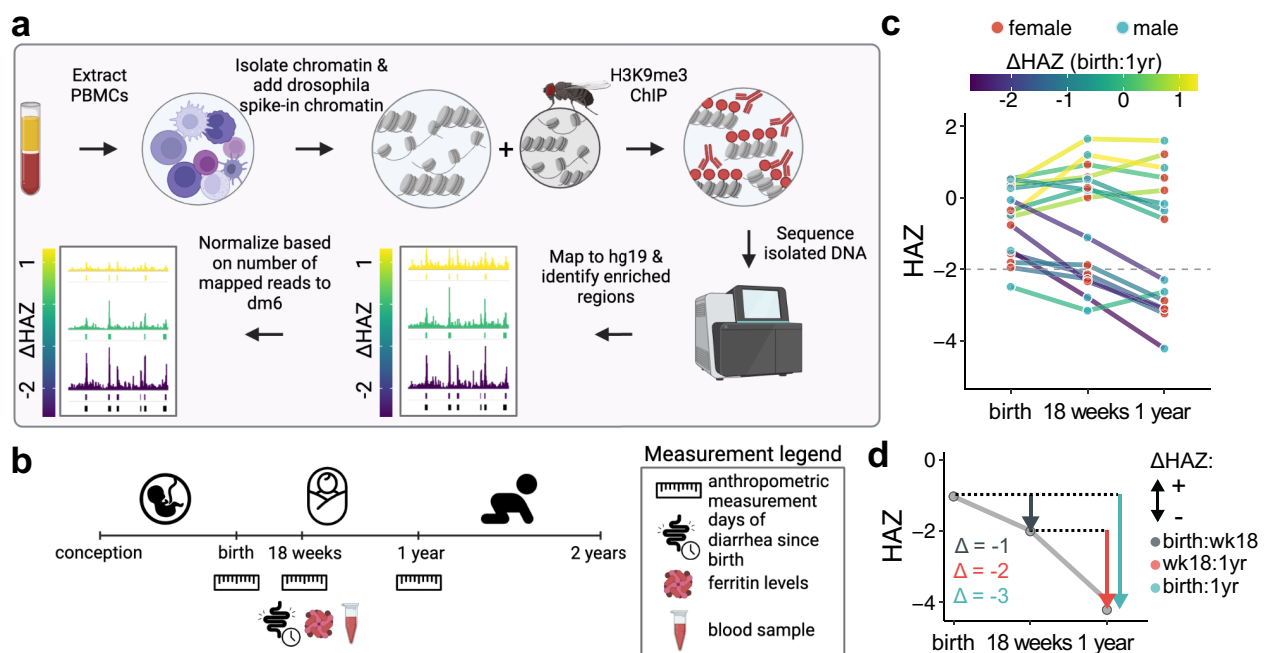


Fig. 1 Project overview and experimental design. **a** Experimental overview. **b** The timeline of early infancy showing sample collection time point and available anthropometry. Blood samples were collected at 18 weeks of age. Additional information used in this study is highlighted in the diagram under the age of a child at the time of information collection. **c** Overview of the growth trajectories of children whose samples were used in this study. Grey points indicate HAZ score (y-axis) at a given age (x-axis), the growth trajectories are colored based on Δ HAZ between birth and one year, and n indicates the number of individuals. Children with HAZ scores at one year below the dashed line are classified as stunted. **d** Δ HAZ score explained. Grey line presents growth trajectory of a single individual and is created by connecting HAZ scores (y-axis) at given ages (x-axis). The colored arrows indicate changes in HAZ score between two time points, i.e., Δ HAZ. Dark grey—change between birth and 18 weeks, red—change between 18 weeks and one year, green—change between birth and one year

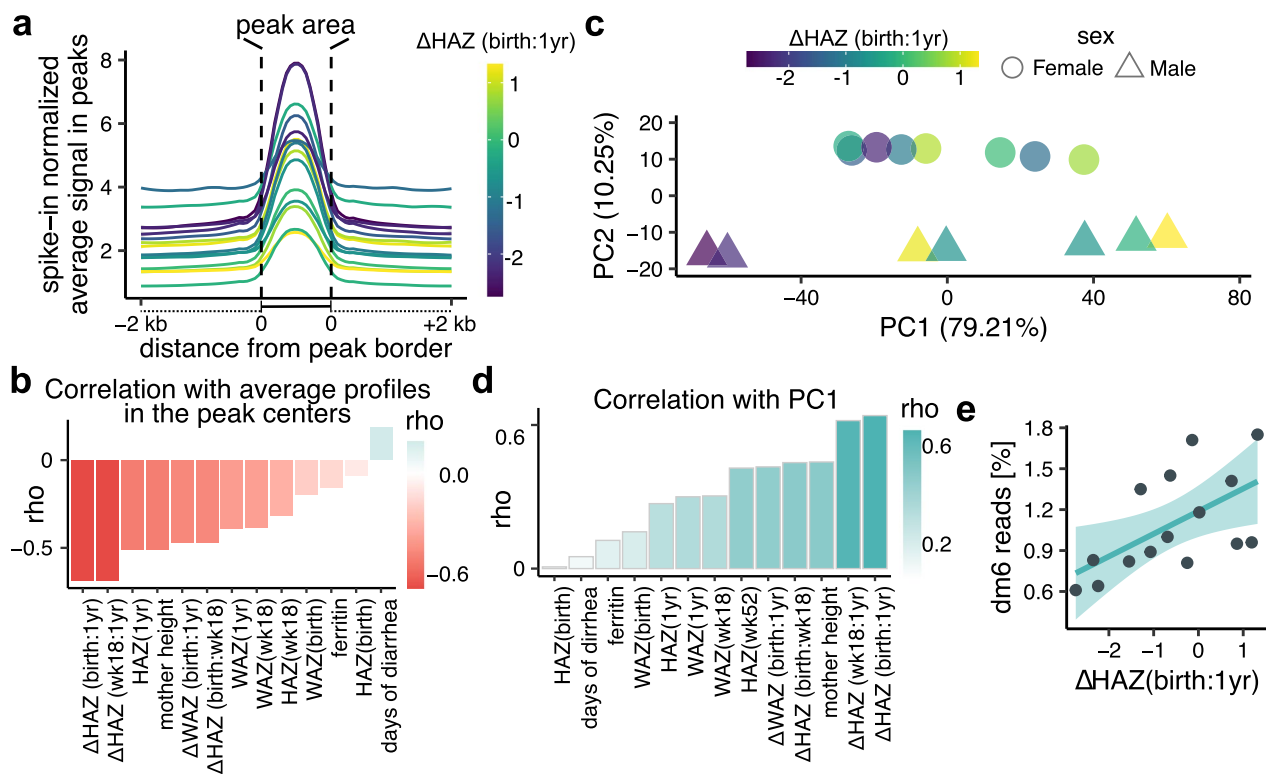


Fig. 2 H3K9me3 profiles are globally correlated with negative growth change between birth and one year. **a** Average H3K9me3 profiles within peaks and 2 kb regions surrounding peaks. Each line represents an average profile of a child color coded by the child's ΔHAZ (birth:1yr) score. The average profiles were normalized by drosophila spike-in normalization factors. **b** Spearman's correlation coefficients (ρ) between average normalized profile values from peak centers in (a) and a given measurement indicated on x axis. **c** PCA plot of H3K9me3 read counts in identified peaks. The numbers in brackets indicate variance explained by a given principal component. **d** Spearman's correlation coefficients (ρ) between PC1 and measurements indicated on x-axis. **e** Linear fit between ΔHAZ (birth:1yr) score and percentage of reads mapped to the *Drosophila melanogaster* genome. The relationship is significant (p -value = 0.021). Degrees of freedom = 13, adjusted R^2 = 0.296

H3K9me3 level, the pattern was much weaker and could have been easily missed without spike-in normalization (Additional file 1: Fig. S5b,c). Furthermore, we show that increased proportional content of *Drosophila* spike-in in the sequencing files, which suggests a genome-wide loss of H3K9me3 (Additional file 1: Fig. S5d), was significantly associated with ΔHAZ (birth:1yr) score (Fig. 2e, Additional file 1: Fig. S5e,f). Altogether, these results suggest that infants undergoing poor growth between birth and one year of age have globally elevated H3K9me3 levels compared to healthier infants.

Differentially affected H3K9me3 regions are associated with stunting relevant pathways

To gain insight into the roles of H3K9me3 regions in stunting, we performed differential analysis in which we asked how H3K9me3 levels changed at each peak with respect to ΔHAZ (birth:1yr) score while controlling for sex. The vast majority of the H3K9me3 regions were significantly upregulated in children with negative changes in HAZ scores during the first year

(Fig. 3a–c, Additional file 2: Table S3), as expected given the significant global upregulation discussed above. The mean fold change per unit decrease of ΔHAZ (birth:1yr) score within significantly affected regions was 1.18 (Additional file 1: Fig. S6a), thus indicating an 18% change in the H3K9me3 signal per unit ΔHAZ (birth:1 yr) score. For functional analysis, we focused on the top 10% of the most significantly affected regions. These regions were not highly specific for a particular cell type (Additional file 1: Fig. S6b), and compared to the H3K9me3 regions overall, they are closer to genes (Fig. 3d), including being notably more enriched in intronic regions (Fig. 3e). The enrichment in intronic over intergenic regions is specific for the top 10% of the significant regions and the enrichment declines rapidly with further ranking (Additional file 1: Fig. S6c,d). We next asked if these regions are enriched with co-localizing histone marks and DNA-binding factors (DBFs) (Additional file 1: Fig. S6e). As expected, we found the highest enrichment was with the known sets of H3K9me3 regions,

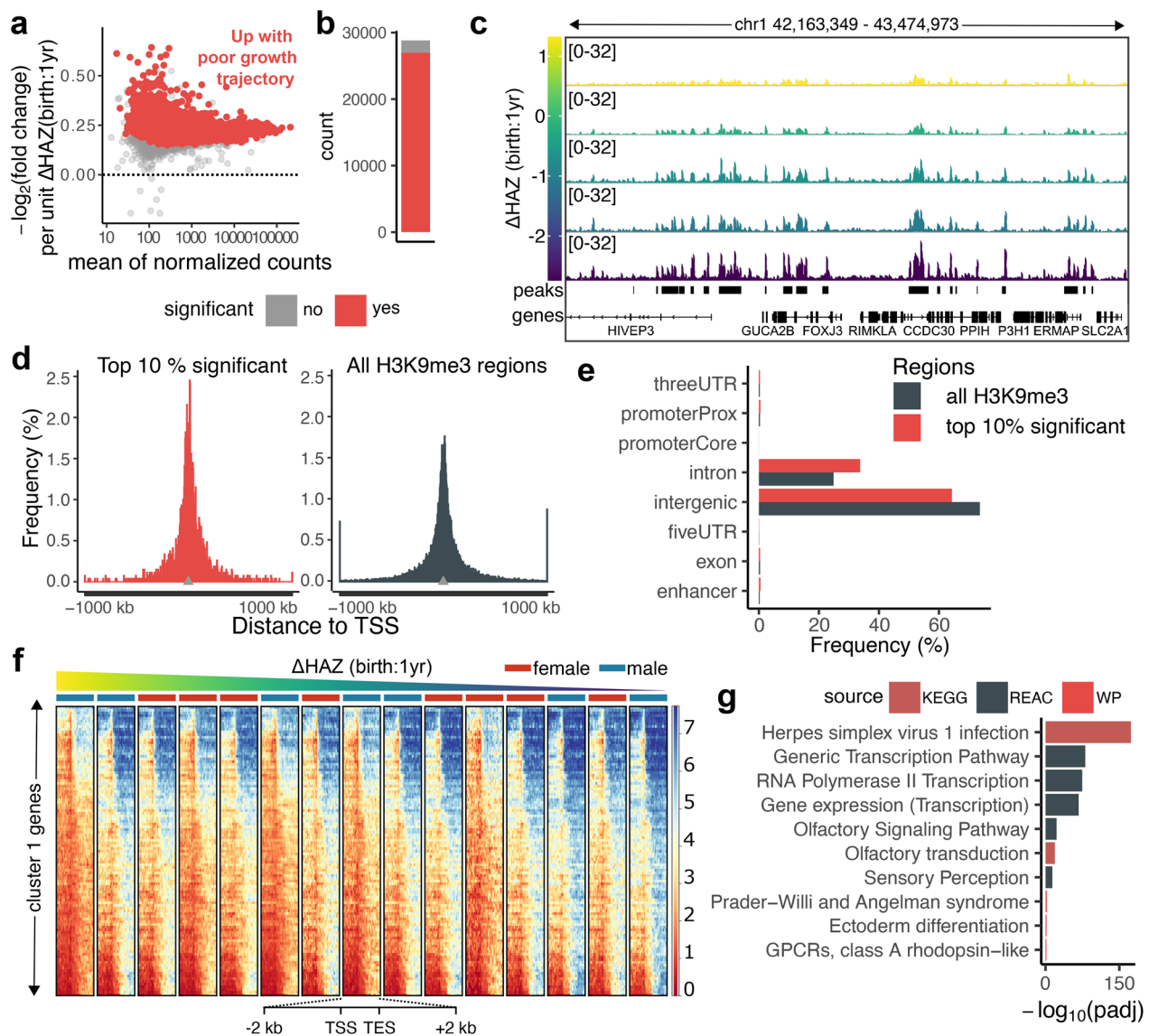


Fig. 3 The most significant regions are nearby gene sites and highly enriched in viral infection pathways. **a** MA-plot shows changes of H3K9me3 levels with negative unit changes of $\Delta\text{HAZ}(\text{birth:1yr})$ score. Each point is an H3K9me3 region, x-axis shows the mean normalized H3K9me3 coverage and y-axis shows $-\log_2(\text{fold change})$ of H3K9me3 coverage per unit increase of $\Delta\text{HAZ}(\text{birth:1yr})$ score. Regions whose coverage increased in association with negative growth trajectory are above the x-axis. Regions with significant association (padj < 0.05) are highlighted in red. **b** Bar plot showing the number of H3K9me3 regions significantly associated vs. not significantly associated with $\Delta\text{HAZ}(\text{birth:1yr})$ score in red and grey respectively. **c** Genome browser snapshot showing an example region, where normalized H3K9me3 coverage increased with poor growth trajectory indicated by low $\Delta\text{HAZ}(\text{birth:1yr})$ score. **d** Histograms comparing the distances of the top 10% most significant (left) vs. all (right) H3K9me3 regions to TSSs. Shown are distances closer than ± 1 Mb from TSSs; more distal regions were accumulated into bins on both sides of the histogram. The distances from TSSs when not accounting for directionality are significantly higher within all H3K9me3 regions compared to the 10% significant regions (two sample t-test, $H_0: \mu_{\text{top10}} = \mu_{\text{all}}$, $H_1: \mu_{\text{top10}} < \mu_{\text{all}}$, $p < 2.2\text{e-}16$). **e** Bar plot showing relative distribution of the top 10% most significant vs. all H3K9me3 regions over genomic classes. The differences in the overlaps between the two classes are significant (Chi-square test, H_0 : the number of overlaps within the top 10% significant and all H3K9me3 regions are independent, $p < 2.2\text{e-}16$). **f** Heatmaps showing normalized H3K9me3 coverage over genes and ± 2 kb surrounding regions from cluster 1 identified in Additional file 1: Fig. S8. Heatmaps are organized based on child's $\Delta\text{HAZ}(\text{birth:1yr})$ score. **g** Results of gene enrichment analysis of cluster 1 genes shown in (f). Plotted are $-\log_{10}(\text{padj})$ of significant enrichments (padj < 0.05). Results come from three different sources: KEGG, Reactome (REACT), and WikiPathways (WP)

but also we found significant enrichment in enhancer regions marked by H3K27ac or H3K4me1 (histone H3 monomethylation on lysine 4) [46], and in H3K36me3 (histone H3 trimethylation on lysine 36). Colocalization of H3K9me3 with H3K36me3 has been shown to be crucial for repressing enhancer activity [47]. Interestingly, we further found significant enrichment in binding sites for the pioneer transcription factor *PU.1* (*SPI1*), a factor crucial for hematopoietic cell fate control [48, 49], suggesting that in stunting, increased H3K9me3 levels at these sites may prevent proper immune cell development through restricting *SPI1/PU.1* binding [50].

The assignment of these top H3K9me3 regions to their gene targets revealed genes that are involved in immune system processes, tissue development, nitrogen and phosphate metabolism and stress responses (Additional file 1: Fig. S7a). Importantly, mouse studies identified many of these gene targets to be involved in processes tightly associated with stunting (Additional file 1: Fig. S7b), such as altered interferon gamma (IFN- γ) levels [51], abnormal enterocyte proliferation associated with environmental enteric dysfunction (EED) [4, 17], or neurological pathways, such as altered post-tetanic potentiation, which potentially impacts neurocognitive development [5, 6].

Genes associated with high H3K9me3 levels are highly enriched in viral immune response pathways

The proximity of these H3K9me3 regions to genes prompted us to explore the H3K9me3 patterns within genes. Two distinct clusters of genes were identified (Additional file 1: Fig. S8a). Cluster 1 contains a smaller subset of genes ($n=1258$) whose coverage changed dynamically with Δ HAZ (birth:1yr) score (Fig. 3f), while the genes in cluster 2 had generally very low H3K9me3 coverage. The cluster 1 genes were enriched in transcription regulation, as well as nucleotide biosynthesis and metabolism, protein deubiquitination, and sensory development (Fig. 3g, Additional file 1: Fig. S8b). The most significantly enriched pathway, however, was Herpes simplex virus 1 infection. We explored the distribution of the genes of interest within this pathway and surprisingly found that 243 out of the 244 genes of interest are genes coding for zinc-finger antiviral proteins (ZAPs) that bind GC-dinucleotide enriched viral mRNA sequences and target them for degradation [52]. If increased H3K9me3 levels indicate repression of these genes, these results could suggest that one of the mechanisms through which H3K9me3 predisposes children to stunting is through paralyzing immune responses to viral invasion.

Elevated H3K9me3 regions appear to mask TEs with potential regulatory roles in immune responses

Since H3K9me3 is well known to localize to and suppress repeated sequences and various classes of transposable elements [37], we next explored the endogenous retroviruses (ERVs) and other classes of transposable elements (TEs) that overlapped with all significantly affected H3K9me3 regions (Additional file 1: Fig. S9a), as well as the top 10% of significant regions (Fig. 4a). In both sets, we identified similar overrepresented classes of TEs, with higher overrepresentation in the top 10% significant regions. We therefore focused on delineating the possible functional relevance of the highly overrepresented TE classes within the top 10% significant regions, specifically the ERVs, long interspersed nuclear elements-L1 (LINE L1), SINE-R-VNTR-*Alu* (SVA) retrotransposons, and telomeric satellite DNA. Overall, we observed highly significant overlap of ERVs (Fig. 4b) and LINE L1 (Additional file 1: Fig. S9b) with regulatory regions for genes implicated in immune response pathways, including cytokine production and immune cell activation, as well as in DNA damage response pathways, catabolic pathways, and cation homeostasis pathways. Interestingly, we also noted an enrichment of H3K9me3 regions overlapping SVA retrotransposons in nutrient sensing pathways (Additional file 1: Fig. S9c), however, with much lower significance compared to the enrichment of pathways relevant to ERVs and LINE L1. Telomeric satellite DNA associated with H3K9me3 was also modestly associated with genes implicated in adhesion (Additional file 1: Fig. S9d).

The results from these analyses suggest that high H3K9me3 levels at these gene regulatory sites suppressed immune responses. We previously reported increased H3K27ac targeting genes associated with viral infection [36]. We therefore tested whether ERVs within the top significant H3K9me3 regions share common gene targets with significantly upregulated H3K27ac regions (Fig. 4c). The majority of gene targets were specific to ERVs or H3K27ac, except for 6 genes that, however, did not show any significant pathway enrichment (Fig. 4d). Expanding the analysis to ERV regions overlapping all significant H3K9me3 regions increased the number of overlapping genes to 61, but even these were not associated with any specific pathways.

Analysis of maternal H3K9me3 profiles

To determine the relationship between maternal and child's H3K9me3 profiles, we performed H3K9me3 ChIP-seq on PBMC samples from mothers ($n=14$, Fig. 5a) enrolled in the same PROVIDE study. Through differential analysis, we first explored which H3K9me3

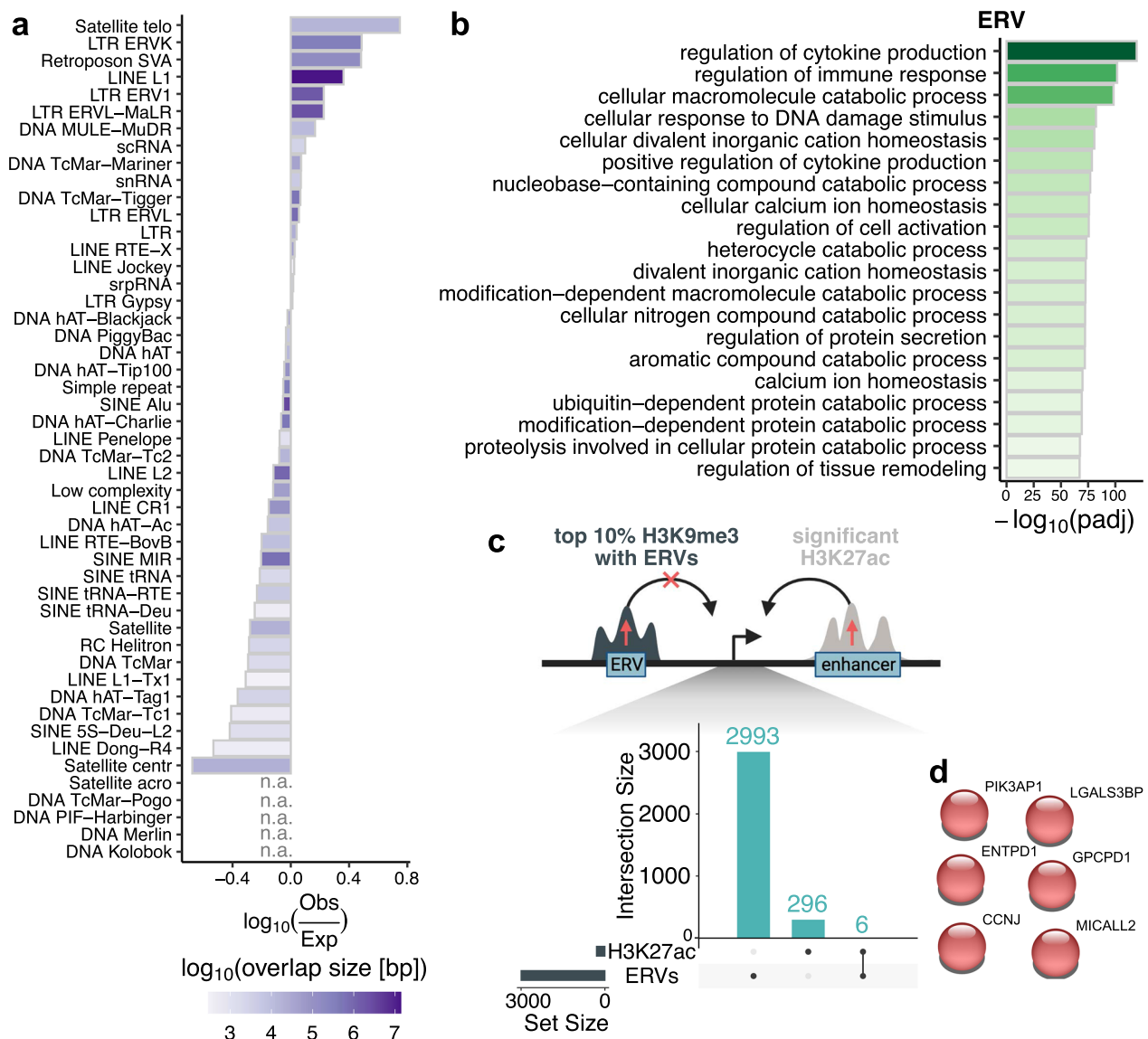


Fig. 4 Misregulated H3K9me3 regions are enriched in transposable elements with potential roles in immune system modulation. **a** Ratio of observed over expected overlaps of TEs with top 10% significantly misregulated H3K9me3 regions with ΔHAZ (birth:1yr) score. Positive values indicate overlaps higher than expected. The color shows the total overlap size in base-pairs for a given TE class. **b** Top 20 GO:BP terms, as displayed by GREAT, associated with ERVs within the top 10% significantly affected H3K9me3 regions. Plotted are negative \log_{10} FDR-corrected p-values (padj). Padj values < 0.05 were considered significant. **c** Comparison of gene targets shared between enhancers with increased H3K27ac values at 18-weeks in stunting [36] and ERVs within the top 10% significant regions with increased H3K9me3 levels as illustrated in the top part of the panel. **d** Shared target genes between H3K27ac and ERVs from **c**. No significant functional enrichment was identified

regions were misregulated in relationship to the child's ΔHAZ (birth:1yr) score while controlling for the child's sex (Additional file 1: Fig. S10a, Additional file 3: Table S4). While we did not observe any significant relationship between maternal H3K9me3 and a child's ΔHAZ (birth:1yr) score, or a strong correlation between the differential analysis results from children and mothers (Fig. 5b), we noticed that the majority of the maternal H3K9me3 regions were upregulated in mothers whose

children had low ΔHAZ (birth:1yr) scores, partially capturing the global pattern observed in 18-week-old infants (Fig. 5c, Additional file 1: Fig. S10b). Further exploration of global relationships between H3K9me3 profiles of mothers and children through correlation analysis of H3K9me3 profiles resulted in identification of two distinct clusters separating mothers from infants (Fig. 5d) suggesting that H3K9me3 carries strong age-related signatures that are possibly masking weaker signatures

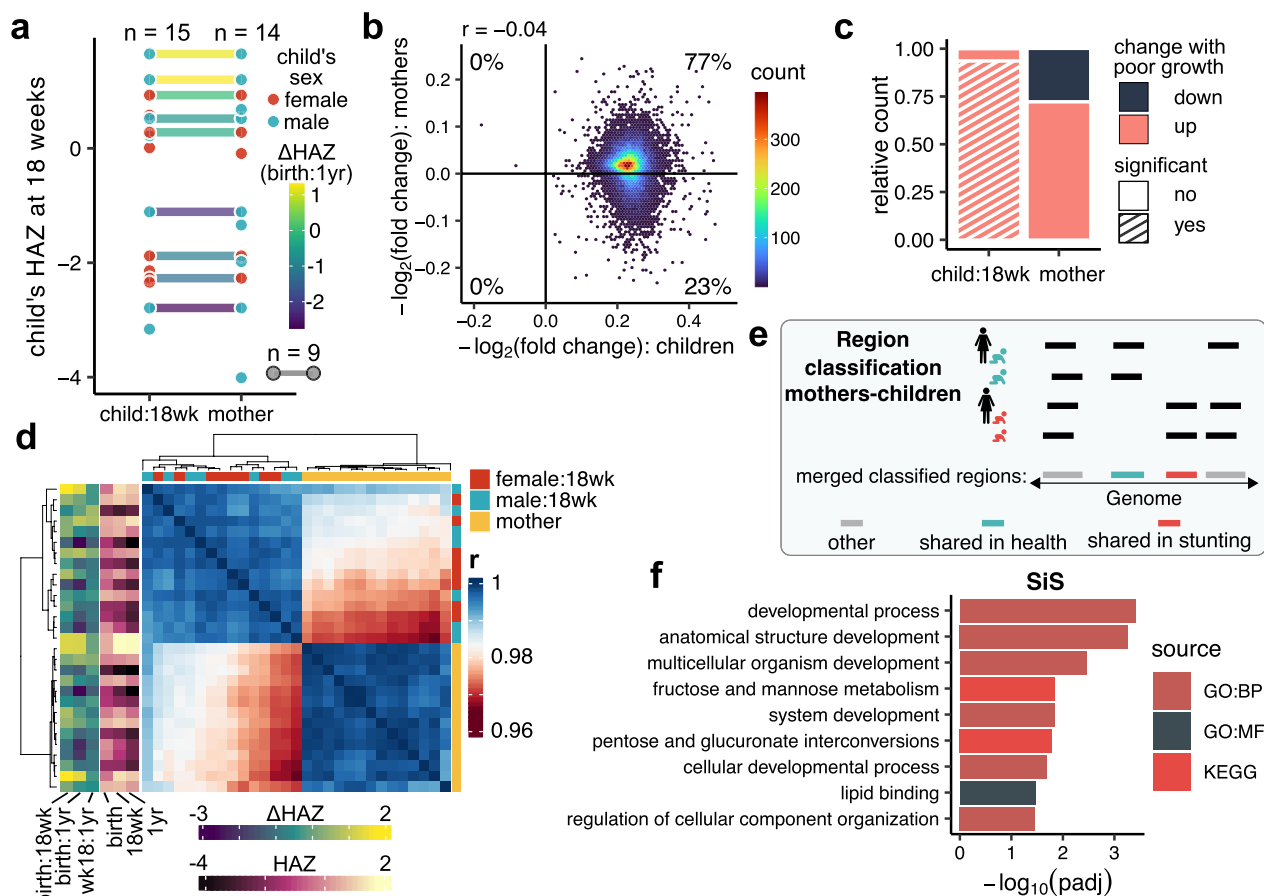


Fig. 5 Relationships between H3K9me3 profiles of mothers and 18-week-old infants. **a** Overview of samples used in this study. Each dot represents an individual whose sample was used in this study. X-axis separates 18-week-old children from mothers, y-axis represents child's HAZ score at 18 weeks of age. A link is drawn only between matching mother-child pairs and is colored based on child's Δ HAZ (birth:1yr) score. The three different *n*s indicate the number of 18-week-old children (left), mothers (right) and matching mother-child pairs (bottom legend). **b** Comparison of differential analysis results showing $-\log_2(\text{fold changes})$ per unit of child's Δ HAZ (birth:1yr) score. On the x-axis are plotted results for 18-week-old children, on the y-axis results for mothers. Each point is an H3K9me3 region identified both in mothers and 18-week-old children. Positive values indicate increasing H3K9me3 levels in association with poor growth trajectory. The color coding shows the density of the points. The numbers indicate the percentage of points falling into a given quadrant. **c** Bar plot summarizing results of differential analysis from 18-week-old children vs. mothers. Regions classified as going up with poor growth trajectory are those with positive $-\log_2(\text{fold changes})$ per unit of child's Δ HAZ (birth:1yr) score and vice versa. **d** Heatmap showing Pearson's correlation coefficients between H3K9me3 profiles of all 18-week-old children and all mothers whose samples were used in this study. **e** Illustration showing classification of regions. Regions identified only in mothers of healthy children (green icon) and in healthy children are classified as SiH (green). Regions identified only in mothers of stunted children (red icon) and in stunted children are classified as SiS (red). **f** Gene set enrichment results for genes associated with SiS regions. Plotted are $-\log_{10}(\text{padj})$ values of significant enrichments ($\text{padj} < 0.05$). Results come from three different sources: KEGG, GO:MF, GO:BP

of stunting. Interestingly, PCA separated maternal and infant samples with PC2, while PC1 still captured 81% of the variance in the data and preserved the strong relationship of the infant H3K9me3 profiles to Δ HAZ (birth:1yr) score (Additional file 1: Fig. S10c,d). The separation of maternal H3K9me3 profiles along PC1 did not reveal a strong correlation with any of the biomarkers (Additional file 1: Fig. S10d). Furthermore, narrowing focus to only the top 10% of the misregulated regions in stunted children did not reveal any new patterns in

comparison to the analysis performed using all H3K9me3 regions (Additional file 1: Fig. S10e,f).

To determine if subtler patterns in the maternal H3K9me3 profiles are potentially contributing to stunting, we tested whether there were regions shared only between mothers of stunted children and stunted children, here termed shared in stunting (SiS). We also tested whether likewise there were maternal H3K9me3 regions in mothers of healthy children that were identified in healthy children (SiH) (Fig. 5e). We identified

110 SiS and 97 SiH H3K9me3 regions (Additional file 1: Fig. S11a) that were not cell-type specific (Additional file 1: Fig. S11b) and did not have a distinct signature in terms of distribution across the genome (Additional file 1: Fig. S11c,d). However, functional annotation revealed a distinct set of pathways specific for each region set. SiS regions were associated with genes involved in multiple developmental processes, lipid binding, and sugar metabolism, specifically fructose and mannose metabolism, and pentose and glucuronate interconversions (Fig. 5f). In contrast, SiH regions were associated with pathways involved in signaling, regulation, development, growth factor regulation, and endothelial cell migration (Additional file 1: Fig. S11e).

Histone marks misregulated in one-year-old stunted children often reside within H3K9me3 regions

To gain more comprehensive insight into epigenetic changes associated with stunting, we combined H3K9me3 results from this study with previously published results describing H3K4me3 [35] and H3K27ac [36] changes associated with stunting in 18-week-old and one-year-old children from the same cohort (but not all the same individuals; Additional file 1: Fig. S12a-d). First, we identified overlaps between H3K9me3 regions and H3K4me3 or H3K27ac regions in both age groups (Fig. 6a). While we observed a decrease in the overlap of H3K9me3 and H3K27ac peaks between 18 weeks and one year, there was an intriguing increase in the overlaps between H3K9me3 and H3K4me3 regions during

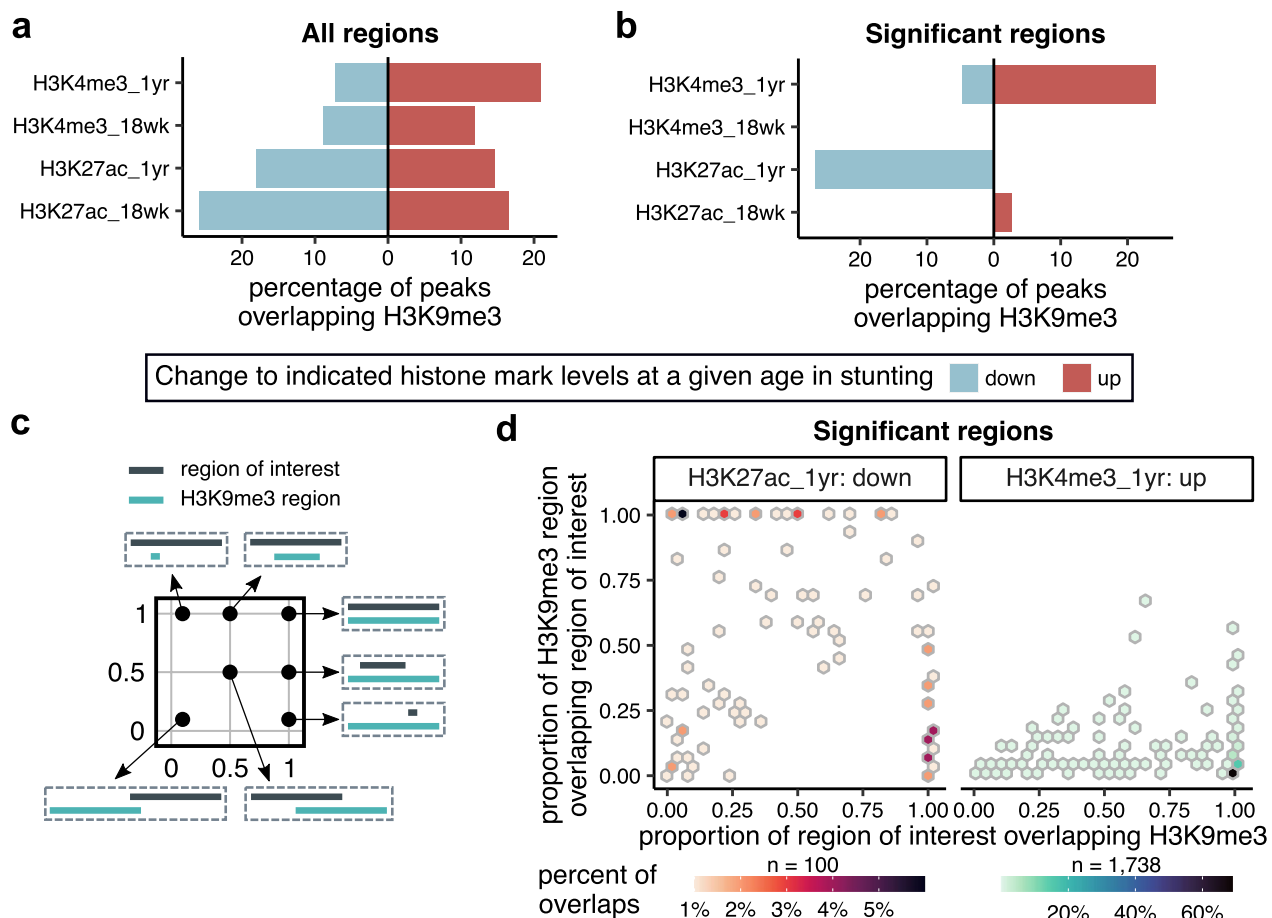


Fig. 6 H3K9me3 regions upregulated in 18-week-old infants overlap active regulatory elements misregulated in 1-year-old stunted children. **a** Bar plot showing percentage of all H3K4me3 and H3K27ac regions identified in 18-week- and one-year-old children that overlap H3K9me3 regions identified in 18-week-old infants. The overlaps were counted separately for upregulated (red) vs. downregulated (blue) regions at a given age in stunting. **b** Bar plot as in **(a)** only showing overlaps with regions that were significantly different in stunted children. **c** Explanatory illustration for panel **(d)** showing different overlap scenarios. **d** The plots show the proportions of H3K9me3 regions that overlap significantly downregulated H3K27ac (left) or significantly upregulated H3K4me3 (right) in one-year-old children (x-axis) versus the proportions of these H3K27ac and H3K4me3 regions that overlap with H3K9me3 (y-axis). Color coding indicates the percentage of regions falling into a given bin. The total number of overlapping regions for each category is indicated above the individual color bars

infancy, especially in the H3K4me3 regions that were upregulated in stunted children. By further narrowing our focus on significantly affected H3K4me3 and H3K27ac regions in stunting, we found that in 18-week-old children, these regions are almost exclusively distinct from the H3K9me3 regions, compared to the significantly affected H3K4me3 and H3K27ac regions in one-year-old children that tended to colocalize with H3K9me3 signal in 18-week-old children (Fig. 6b, Additional file 1: Fig. S12e-h). Notably, there was an overlap of almost 25% of the small upregulated H3K4me3 regions that were previously suggested to be redistributed from the larger H3K4me3 regions at TSSs [35], and nearly 30% of the H3K27ac regions downregulated in one-year-old stunted children overlapped with H3K9me3 peaks. The downregulated H3K27ac regions in one-year-old stunted children did not have a very specific location relatively to the overlapping H3K9me3 regions, while the upregulated H3K4me3 regions in one-year-old children were predominantly encompassed by H3K9me3 regions (Fig. 6c,d, see Additional file 1: Fig. S13a,b for overlaps with all H3K4me3 and H3K27ac regions). These observations suggest a possible mechanism in which the methylation machinery compensates for the high levels of H3K9me3 observed in 18-week-old infants through increased H3K4me3 levels at one year of age.

Functional annotation of overlaps with H3K4me3 regions upregulated in stunted children at one year suggests that these are involved in regulating transcription, cell adhesion and developmental processes (Additional file 1: Fig. S14a), and there was also a relatively weak colocalization of these regions with H3K36me3 (Additional file 1: Fig. S14b). In contrast, the overlaps with H3K27ac peaks that were downregulated in stunting at one year showed enrichment only in cell activation pathways (Additional file 1: Fig. S14c), in addition to strong colocalization with both activating and repressing histone marks (Additional file 1: Fig. S14d), as well as with a broad range of DBFs (Additional file 1: Fig. S14e).

Discussion

This study was motivated by two goals. The first goal was to uncover epigenetic changes in early infancy associated with the development of stunting. This goal arose from our previously published results in which we observed large-scale stunting-associated changes to H3K4me3 [35] and H3K27ac [36] in one-year-old children suggesting immune system exhaustion and alterations to one carbon-metabolism. However, in those studies we saw no significant changes to H3K4me3 in 18-week-old infants and only relatively smaller changes to H3K27ac which suggested activation of stress responses and genes involved in viral infection. We hypothesized that if there

were epigenetic changes in early infancy predisposing children to become stunted, we would need to investigate a more stable mark associated with development, such as H3K9me3 [53]. The second goal was to determine if there were relationships between maternal and child H3K9me3 profiles that could shed light on the propagation of stunting, including transgenerational inheritance, which contributes to the development of stunting via H3K9me3 and other factors in animal models [42].

The results presented here along with prior work on H3K4me3 and H3K27ac are summarized in Fig. 7. The major finding here is that globally elevated H3K9me3 levels in 18-week-old infants are associated with poor growth between birth and one year of age as indicated by the Δ HAZ (birth:1yr) score (Fig. 2). Notably, H3K9me3 levels show much weaker association with a child's current health status in terms of the HAZ score at 18 weeks (Fig. 2), suggesting that altered H3K9me3 levels do not reflect immediate changes in a child's environment, but instead may play a role in programming developmental pathways over time. Overall, these results suggest the possibility that high H3K9me3 levels could be used as a biomarker for poor growth before the overt manifestation of stunting. We attempted to confirm the changes in H3K9me3 by Western blotting but found that there is a high degree of variability in the values measured by this method relative to the change in H3K9me3 detected by spike-in normalization (Additional file 1: Fig. S4b-d).

Through the detailed analyses presented here, we explored molecular mechanisms through which elevated H3K9me3 levels may predispose children to poor growth. Genes residing within misregulated H3K9me3 regions were involved in Herpes simplex virus 1 infection pathway (Fig. 3f,g). These include a large number of ZAP genes which encode proteins that target viral mRNA for degradation [51]. Thus, their possible repression resulting from elevated H3K9me3 potentially predisposes children to be more vulnerable to viral infection. TEs within the top significant H3K9me3 regions may also contribute to H3K9me3 regulatory mechanisms. While the functional role of TEs is largely unexplored, it has been thought that TEs, such as ERVs, are marked by H3K9me3 to ensure they remain transcriptionally silent. Recent studies also show that TEs can act as binding sites for transcription factors (TFs) [54] and furthermore that they have roles in development of innate immune responses [39]. Here we show that TEs residing within the top significant H3K9me3 regions are in proximity to genes involved in immune system activation, where we observed an especially high enrichment in cytokine production pathways, as well as DNA damage pathways, catabolic pathways, and other pathways with less significant enrichment (Fig. 4b, Fig S9b-d). It has been recently shown that

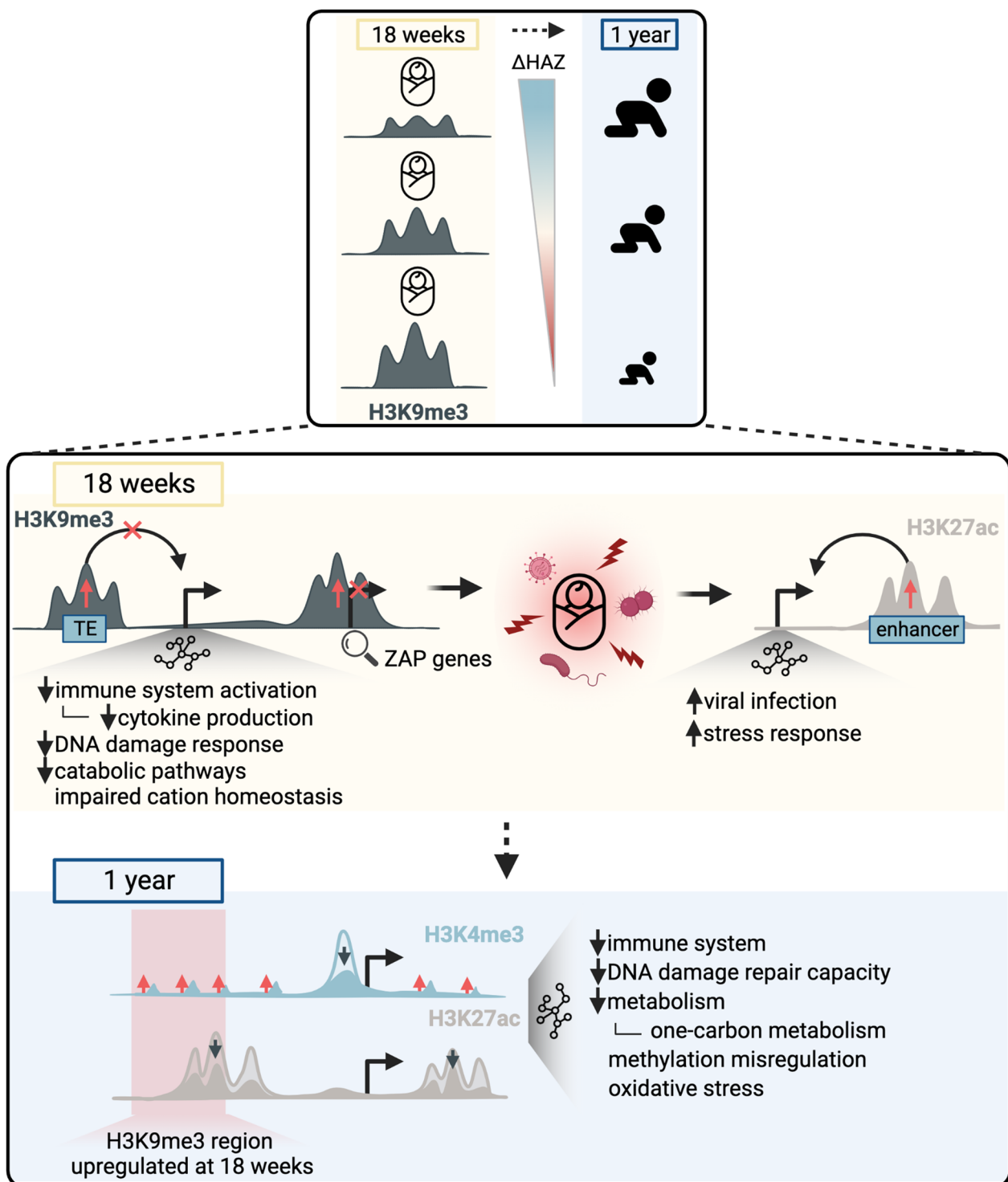


Fig. 7 Summary and working model. Top panel shows that globally increased H3K9me3 levels at 18 weeks of age are associated with poor growth through the model illustrated in the bottom panel. In this model, increased H3K9me3 levels restrict expression of ZAP genes and regulatory activity of key TEs resulting in immature immune responses, and thus high pathogen load and overall stress in children destined to be stunted. These changes then trigger increased stress responses and activation of pathways involved in viral infection as indicated by elevated H3K27ac. As an adaptation to long-term stresses, H3K4me3 and H3K27ac undergo large-scale changes, as described previously [35, 36], which affect the highlighted pathways by one year of age. A significant proportion of H3K4me3 and H3K27ac changes observed in one-year-old children fall into misregulated H3K9me3 regions identified in 18-week-old children

non-repressed ERVs can create competition for TF binding, and in addition, decreased heterochromatin levels at ERVs were associated with overall decreased H3K27ac in other genomic regions [41]. This model fits well with our data in which increased H3K9me3 at ERV sites and increased H3K27ac levels at other sites in the genome suggest deleterious masking of ERV sites. Taken together, these results suggest a model (Fig. 7) in which increased H3K9me3 levels at key TEs and *ZAP* genes may paralyze activation of immune system responses in early infancy, leading to an increased pathogen load and overall stress in children destined to become stunted. This is consistent with the apparent activation of genes associated with viral infection and stress response pathways through elevated H3K27ac.

Our results show that high levels of H3K9me3 in 18-week-old infants are associated with poor Δ HAZ (birth:1 yr) growth score and potentially hamper cytokine production. A recent study performed using samples from the same cohort shows that by one year of age stunting is inversely associated with high cytokine production, specifically of interleukin 8 (IL8) and transforming growth factor- β (TGF β) [55]. The combined results are consistent with the emerging understanding that changes to immune system function are highly dynamic during early development [55], and further suggest that the development of a healthy immune system may involve precise timing in the developmental capacities of specific immune responses in response to pathogen burden.

We also discovered that the top differentially affected H3K9me3 regions are enriched in *SPI1/PU.1* TF binding sites (Additional file 1: Fig. S6e). *SPI1/PU.1* is a TF crucial for immune cell differentiation [48, 49] and for this reason we posit that potential H3K9me3-mediated repression of these sites leads to immune cell immaturity in early infancy. The gene targets of the top differentially affected H3K9me3 regions were further identified in pathways associated with altered circulating IFN- γ levels. Interestingly, these have been previously associated with stunting and susceptibility to amebiasis [51].

Integrated analysis with H3K4me3 [35] and H3K27ac [36] results from children from the same cohort (but not all the same individuals) further showed that, as expected, there was not much of overlap between the H3K9me3 regions and the activating H3K4me3 and H3K27ac marks at 18 weeks of age. However, a relatively high proportion of significantly misregulated H3K4me3 and H3K27ac regions in one-year-old stunted children overlap with the upregulated H3K9me3 regions at 18-weeks (Fig. 6b). About 25% of distal ectopic H3K4me3 regions occur within H3K9me3 regions and almost 30% of the upregulated H3K27ac regions overlap with H3K9me3 regions. The misregulated regions were previously described

by gene association analysis to regulate processes highlighted in Fig. 7. These findings suggest a possible model in which elevated H3K9me3 levels in 18-week-old children preset a detrimental chromatin environment and that predisposes stunted children to mislocalization of H3K4me3 and misregulation of H3K27ac by one year of age. The H3K4me3 pattern is perhaps indicative of a compensating mechanism for previous over-suppression of these important regulatory regions.

We did not identify any significant H3K9me3 regions in mothers with a strong relationship to the child's Δ HAZ (birth:1yr) score. However, the overall directionality of the H3K9me3 changes was in concordance with changes observed in the 18-week-old infants (Fig. 5b,c, Additional file 1: Fig. S10b). This suggests that there may be a relationship between the maternal and child profiles, but one that would probably require a much larger sample size to detect. Additional insight would likely come from analysis of paternal data, however, at the moment we are limited in both the number and types of samples available from this highly vulnerable population.

Overall, this study illustrates the power of epigenomic analyses to provide insight into the etiology of stunting, even when the sample numbers are limited and there are many contributing factors to its development.

Conclusions

Analysis of the heterochromatin-associated histone mark H3K9me3 in PBMCs of 18-week-old infants showed that globally elevated H3K9me3 levels are associated with poor growth between birth and one year. Importantly, these changes were detectable before the overt appearance of the stunted phenotype and therefore present an opportunity for future work aimed at understanding mechanisms that may be manipulated to avert stunting as well as development of new biomarkers to detect it early. Moreover, these results suggest a relationship between H3K9me3 and the immune system dysfunction that develops in stunted children.

Methods

H3K9me3 ChIP-seq of human PBMCs

Deidentified PBMC samples from 18-week-old children ($n=15$) and mothers ($n=14$) enrolled in the PROVIDE study ($n=700$) [44] were obtained in collaboration with icddr,b (International Centre for Diarrhoeal Disease Research, Bangladesh) in Dhaka, Bangladesh. The number of samples was established based on previous work [35, 36], and samples were selected to represent the PROVIDE spectrum of HAZ and Δ HAZ scores, as well as to be balanced in sex representation (Additional file 1: Fig. S1,2). ChIP-seq was performed as previously described in detail in [35]. Briefly, PBMC samples were fixed with

formaldehyde, chromatin isolated and sheared, 0.02% (microgram/microgram) *Drosophila melanogaster* chromatin (cat #53083, Active Motif, Carlsbad, CA, USA) was added for spike-in normalization [56]. H3K9me3 antibodies (12.5 µl per 100 µg chromatin protein solution, Abcam Cat# AB8898, RRID: AB_30684) were used for DNA fragment selection by immunoprecipitation. Sequencing libraries were constructed using the Illumina TruSeq ChIP Library Preparation Kit following the manufacturer instructions. Libraries were sequenced using an Illumina NextSeq500 instrument with high-capacity cartridge in the University of Virginia Genome Analysis and Technology Core, RRID:SCR_018883, yielding between 52 to 142 million 150 bp single-end reads per sample (Additional file 1: Table S1).

Western blotting

Western blotting was performed as previously described in [35]. 5 µg of protein from 18 weeks of age was separated on a 4–20% gradient gel and processed for immunoblotting, using anti-H3K9me3 (Abcam Cat# AB8898, RRID: AB_30684) and then stripped and reprobed for total H3 c-terminus (Active Motif Cat# 39451, RRID: AB_2793242). Secondary: ECL Plex anti-rabbit IgG-Cy5 (Cytiva Cat# PA45011, RRID: AB_772205). Blots were imaged on Typhoon Cy5 600 PMT, Red (633) laser, 670 BP 30 filter.

H3K9me3 ChIP-seq data preprocessing

Bowtie2 (2.2.6) [57] with default settings was used to map raw FASTQ files to the hg19 reference genome obtained with refgenie (0.9.3) [58]. SAMtools (0.1.19-44428cd) [59] were used to convert alignment SAM files to BAM format (function: *view -S -b*), sort (function: *sort*), and index (function: *index*) BAM files, report number of mapped (function: *view -c -F 4*) and unmapped (function: *view -c -f 4*) reads and to filter out unmapped reads (function: *view -h -F 4 -b*). ENCODE blacklisted sites [60] were removed with bedtools (v2.26.0) [61] *intersect* function with argument *-v*. The quality of both raw FASTQ files and final BAM files was checked with FastQC (v0.11.5) [62] together with multiQC (1.11) [63]. To further visually inspect the quality of the data, BAM alignment files were converted to bigWig files (initially without normalizing) and browsed with Integrative Genomics Viewer (IGV_2.7.2) [64]. To generate bigWig files, BAM files were first converted to BED files with coverage information using bedtools *bamtobed*, these files were then converted to COV files using bedtools *genomecov*, and finally bigWig files were generated from the COV files with UCSC *bedGraphToBigWig* (v4) [65]. Chromosome sizes for bedtools *genomecov* and *bedGraphToBigWig* were obtained with refgenie (0.9.3) [58].

Regions of enrichment, i.e., peaks, were generated from BED coverage files using SICER (v1.1) [66] against input (separate input files for children and mothers) with *redundancy threshold*=1, *window size*=1000, *fragment size*=425, *effective genome fraction*=0.74, *gap size*=5000, and *FDR*=0.01.

Count tables with the number of mapped reads for each identified region were created for children and mothers separately. Identified peaks were first merged (mothers and children separately) with bedtools *merge* and quantification was done with bedtools *multicov* using BAM alignment files and merged regions.

Code is available from: https://github.com/AubleLab/H3K9me3_stunting.

Normalization

Drosophila spike-in reads were used for normalization to account for possible global changes in H3K9me3 levels (Additional file 1: Fig. S5a). To retrieve the number of spike-in reads within each sample, raw FASTQ files were mapped to the *Drosophila* dm6 reference genome (obtained with refgenie) using bowtie2 with default parameters. SAMtools (0.1.19-44428cd) [59] were used to convert alignment reads to BAM format (function: *view -S -b*), sort (function: *sort*), and index (function: *index*) BAM files, ENCODE blacklisted sites [60] were removed with bedtools (v2.26.0) [61] *intersect* function with argument *-v*, and finally SAMtools were used to report number of mapped reads (function: *view -c -F 4*). The final normalization factors were then determined by dividing the number of reads that mapped to dm6 by an arbitrary constant (500,000). These normalization factors were then used in differential peak analysis with DESeq2 (1.30.1) [67]. Inverted values of the normalization factors (1/NF) were used to generate normalized bigWig files, specifically bedtools *genomecov* function with argument *-scale* set to the inverted normalization factor was used to create normalized bedGraph files from BAM alignment files, which were further converted to normalized bigWig files with UCSC *wigToBigWig* tool (v4) [65].

Correlations between the percentage of reads mapped to the dm6 reference genome (indicative of globally misregulated H3K9me3 levels, see Additional file 1: Fig. S5d) with all available biomarkers/measurements were calculated using R *cor* function with *method* = “*spearman*”.

The relationship between the percentage of reads mapped to dm6 and a child's ΔHAZ (birth:1 yr) score was determined by fitting a linear model with the R *lm* function, where the p-value < 0.05 was considered significant.

Differential analysis

DESeq2 (1.30.1) [67] was used to identify differentially regulated H3K9me3 regions in both child and maternal

datasets. Raw count tables along with drosophila normalization factors were passed to DESeq2 with a formula set to child's sex + child's Δ HAZ (birth:1 yr) score. Regions with FDR corrected p-values (padj) less than 0.05 were identified as significantly different. Other combinations of formulas combining a child's sex with other available biomarkers were tested (data not shown), however, no significant regions were identified.

PCA

PCA was performed using *prcomp* R function applied to spike-in- and regularized logarithm- (DESeq2 (1.30.1) [67] *rlog* function) normalized count tables. Counts from all available regions were used for PCA. Results were plotted using *ggplot2* (3.3.6) [68] and points colored based on child's Δ HAZ (birth:1 yr) score with a palette from *viridis* package (0.6.2) [69]. PCA performed on both children's and maternal data together was applied on spike-in-normalized counts over merged (bedtools *merge* function) children's and maternal H3K9me3 regions. Correlations between PC1 with all available biomarkers/measurements were calculated using R *cor* function with *method* = "spearman".

Average profiles

Average spike-in-normalized H3K9me3 profiles were obtained by first using the deepTools (3.3.1) [70] *computeMatrix* function to quantify spike-in-normalized bigWig files over H3K9me3 regions identified in children's datasets. The *computeMatrix* parameters were set to *scale-regions* mode with *-a 2000*, and *-b 2000* parameters to expand average profiles by additional ± 2 kb surrounding the H3K9me3 regions. From *computeMatrix* output, the values for average profiles were obtained with the deepTools *plotProfile* function with the *outFileNameData* parameter included, to get the actual average profile values in addition to the figure output. The resulting average profile values were then plotted using *ggplot2* (3.3.6) [68], where smoothing of the profiles was done with *geom_smooth* function with parameters set to *method* = "loess", *span* = 0.2 and colored based on child's Δ HAZ (birth:1 yr) score using *viridis* package (0.6.2) [69].

Average profiles using read count per million mapped reads normalization were obtained by mapping BAM alignment files onto the H3K9me3 regions with *ngs.plot* (2.61) [71] with default settings, which automatically returns average profiles normalized to read count per million mapped reads. These profiles were then replotted with *ggplot2* to include coloring based on a child's Δ HAZ (birth:1 yr) score with *viridis* package.

Correlations between the centers of average profiles with all available biomarkers/measurements were calculated using the R *cor* function with *method* = "spearman".

Functional annotation of children's H3K9me3 regions

The top 10% of H3K9me3 regions (based on FDR-corrected p-value, followed by p-value) that were identified in children's datasets as differentially misregulated versus child's Δ HAZ (birth:1 yr) score were compared with all of the H3K9me3 regions identified in the children's datasets using GenomicDistributions (1.5.5) [72]. *CalcFeature-DistRefTSS* function with default parameters was used to compare the distances of the regions to the nearest TSSs (transcription start sites), and the distances were plotted with *plotFeatureDist* function with parameters *feature-Name* = "TSS", *size* = 1000000, *infBins* = *T*, *nbins* = 300. *CalcPartitions* function was used to identify overlaps with genomic classes. The genomic classes used as an input to the function were the default classes provided by the GenomicDistributions package with PBMC enhancer regions added from EnhancerAtlas 2 [73]. To create a list of PBMC enhancer regions, enhancer regions from relevant cell types were downloaded from EnhancerAtlas 2, namely CD4+, CD8+, CD14+, CD19+, CD20+, GM10847, GM12878, GM12891, GM12892, GM18505, GM18526, GM18951, GM19099, GM19193, GM19238, GM19239, GM19240, and PBMC cells. Finally, the regions were merged using the bedtools (v2.26.0) [61] *merge* function. The parameters to the *calcPartitions* function were set to *bpProportion* = *T*, and the overlaps were plotted with *plotPartitions* function with default parameters. The cell-type specificity of the regions was determined by using *calcSummarySignal* function with provided H3K9me3 cell-type specific signal matrix (see "Normalized cell-type specific H3K9me3 signal matrix" section of Methods) and plotted with *plotSummarySignal* function with parameters *plotType* = "violinPlot", *metadata* = *cellTypeMetadata*, *colorColumn* = "tissueOfOrigin". A modified version of *calcCumulativePartitions* function was used to calculate cumulative distribution of the overlaps with genomic classes. The original function sorts regions based on their size; the modified function accepts presorted regions. The genomic classes used were identical with the ones used in *calcPartitions* function. The log₂(fold change) of cumulative distribution values of intergenic/intronic regions were calculated and smoothed for plotting using R *loess* function with parameter *span* = 0.15.

Histone mark and DBF (DNA-binding factor) enrichment of the top 10% significantly misregulated regions in children was calculated using LOLA (1.20.0) [74] against the CISTROME [75, 76] database of histone marks and transcription factors (here referred to as DBFs) filtered for blood cell types. The *userUniverse* parameter in LOLA was set to all defined children's H3K9me3 regions, otherwise all default parameters were used. Significant enrichments were defined as those with q-values < 0.05.

The most significant cell-type/antibody combinations were plotted in form of a dot plot, where dot sizes and dot colors reflect the significance of the enrichment for a given cell-type/antibody combination. Note that CIS-TROME provides only regions for the hg38 genome assembly, therefore all H3K9me3 regions used here were first converted to hg38 using UCSC *liftOver* tool [65].

Gene set enrichment of the top 10% significantly mis-regulated regions in children was done with GREAT [77]. Results were then reported as bar plots. Due to a larger number of identified GO:BP (Gene Ontology: Biological Process) terms, those were not reported in the bar plot and instead were first processed with REVIGO [78] with default parameters. The resulting network was further reorganized in Cytoscape (3.8.0) [79] and terms were manually classified into groups.

Normalized cell-type specific H3K9me3 signal matrix

A matrix containing normalized H3K9me3 values across different cell types was used to infer cell-type specificity of regions of interest with GenomicDistributions (1.5.5) [72] *calcSummarySignal* function. To create the H3K9me3 cell-specific matrix, ENCODE [80] data was used. First, BAM alignment files from all available H3K9me3 ChIP-seq experiments on primary cells were used (ENCODE filters: Assay title—Histone ChIP-seq/Target of assay—H3K9me3/Organism—*Homo sapiens*/Biosample classification—primary cell/Genome assembly—hg19/Available file types—BAM). The downloaded BAM files were first sorted and indexed with SAMtools (0.1.19-44428cd) [59] *sort* and *index* functions, respectively. To define H3K9me3 regions across genome, BED files were also obtained from ENCODE (ENCODE filters: Assay title—Histone ChIP-seq/Target of assay—H3K9me3/Organism—*Homo sapiens*/Biosample classification—primary cell/Genome assembly—hg19/Available file types—BED narrowPeak). BED files were sorted (*sort -k1,1 -k2,2n*) and merged with bedtools (v2.26.0) [61] *merge* function. Raw signal values were quantified by using bedtools *multicov* function on H3K9me3 processed BAM files with merged H3K9me3 BED file. The resulting raw count tables were further processed in R. First, samples from the same cell-type origin were merged by summing the cell-type replicate signal values, followed by normalization of the signal values for individual cell types. The normalization was performed in three steps: i) signal values above the 99th percentile for a given cell type were set to 1; ii) signal values below the 99th percentile for a given cell type were normalized to range between 0 and 1 using the following formula: $n_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$, where n_i is the normalized signal value for region i , x_i is the raw signal value for a region i , and x is a vector of raw signal values for a given cell type;

and iii) signal values ranging between 0 and 1 within a given cell type were further quantile-normalized with *normalize.quantiles* function from preprocessCore package (1.52.1) [81].

Gene coverage heatmap

Heatmaps with spike-in-normalized coverage over genes were generated using deepTools (3.3.1) [70] *computeMatrix* function with *scale-regions* mode, and parameters *—sortRegions descend, -b 2000, -a 2000*. Mapped were spike-in-normalized bigWig files onto a BED file containing hg19 coordinates of protein-coding genes along with information about strandness. The gene coordinates were obtained with EnsDb.Hsapiens.v75 (2.99.0) [82], ensemblDb (2.14.1) [83], and AnnotationFilter (1.14.0) [84] packages, where ensemblDb *genes* function was applied to EnsDb.Hsapiens.v75 object with *AnnotationFilter* set up to *~gene_biotype = "protein_coding"*. Only standard chromosomes were kept using GenomeInfoDb (1.26.7) [85] *keepStandardChromosomes* function. The output from the *computeMatrix* function was then plotted as a heatmap with deepTools *plotHeatmap* function with parameters *—kmeans 2 —whatToShow 'heatmap and colorbar'*.

Genes reported in cluster 1 were then passed to g:Profiler [86] for gene set enrichment analysis. Only KEGG, Reactome, WikiPathways and GO:BP terms were searched, as suggested by Reimand et al. [87]. Significant terms ($\text{padj} < 0.05$) were plotted as bar plots showing $-\log_{10}(\text{padj})$ values, except for GO:BP terms (due to a larger number of identified GO:BP terms), which were first processed by REVIGO [78] with default parameters, resulting network was further reorganized in Cytoscape (3.8.0) [79] and terms were manually classified into groups.

Analysis of transposable elements

A library of TEs (transposable elements) was obtained from RepeatMasker [88] library 20140131 for the hg19 genome assembly. Genomic coordinates for individual TE classes were extracted into BED files using R. Unambiguous classes such as RNA, DNA, or classes with a question mark in them were removed from further analysis. Overlaps with H3K9me3 regions were calculated using GenomicDistributions (1.5.5) [72] *calcExpectedPartitions* function with arguments *bpProportion = T*, and *genomeSize* set to sum of hg19 chromosome sizes, as instructed in the GenomicDistributions manual. Overlaps were then plotted with the *plotExpectedPartitions* function with added ggplot2 layer to show color intensity based on the overall size of overlap in number of base pairs.

To functionally annotate interesting TEs, intersections of all ERVs, LINE L1, SVA retrotransposons, and

telomeric satellite DNAs with the top 10% significant, as well as all defined H3K9me3 regions, were obtained with the bedtools (v2.26.0) [61] *intersect* function. Intersections with the top 10% H3K9me3 regions were passed to GREAT [77], where intersections with all H3K9me3 regions were used as background. Maximum top 20 GO:BP terms, as displayed by GREAT, were then plotted with ggplot2 (3.3.6) [68] as bar plots showing negative \log_{10} of FDR-corrected p-values.

Gene targets with significantly increased H3K27ac in 18-week-old children were obtained from Kupkova et al., 2021 [36] and compared to the gene targets of all ERVs within the top 10% significant H3K9me3 regions annotated by GREAT. Results were plotted as an upset plot using the *upset* function from the UpSetR package (1.4.0) [89]. Gene targets shared between ERVs and H3K27ac regions were visualized and functionally annotated with STRING [90].

Relationships between H3K9me3 in mothers and children

PCA applied to both sets of data was performed as described in the PCA section of Methods.

Results from differential analysis (negative \log_2 (fold change) per unit change of child's Δ HAZ score for both children and mothers) were plotted as box plots using ggplot2 (3.3.6) [68] and two-sided one-sample t-tests were used to determine any global unidirectional changes ($H_0: \mu_{\log_2FC} = 0$, $H_1: \mu_{\log_2FC} \neq 0$, $\alpha = 0.05$). To show direct comparisons of \log_2 (fold changes) between children and mothers, H3K9me3 regions from mothers were overlapped with H3K9me3 regions from children with the bedtools (v2.26.0) [61] *intersect* function with parameter *-wao*, which returns the original coordinates for overlapping regions. \log_2 (fold changes) from the overlapping regions were then plotted as a scatter plot with ggplot2 and the Pearson's correlation coefficient was calculated with the R *cor* function.

Correlation analysis between maternal and children's samples was performed by first merging all maternal and child regions with bedtools *merge* followed by quantification of all maternal and child samples with bedtools *multicov*. A correlation matrix was then created by passing the count table to the R *cor* function with *method* = "pearson" and plotted as a heatmap using the *Heatmap* function from ComplexHeatmap (2.6.2) [91], where both rows and columns were clustered using average linkage.

To identify regions specific to mothers of stunted children/mothers of healthy children/stunted children/healthy children, all regions from all child and maternal datasets were first merged with the bedtools *merge* function to create a list of "master regions". With the same function, regions of individuals belonging to a given category were merged, creating "category regions". To

compare the regions shared among or specific to categories, overlaps were found between "master regions" and individual "category regions" using bedtools *intersect* with the *-wao* parameter, which reports the original entries rather than the intersections. "Master regions" that were identified as overlapping with a given category were then passed to the *upset* function from UpSetR package (1.4.0) [89] to reveal the number of regions shared between a given combination of categories. Regions "shared in stunting" were defined as those "master regions" that had overlaps found with regions from mothers of stunted children and with regions from stunted children, but no other category, and analogically were defined regions "shared in health". Summarization of the regions shared in health and shared in control individuals was performed using GenomicDistributions (1.5.5) [72], specifically, the *calcFeatureDistRefTSS* function was used to compare the distances to the nearest TSS. The \log_{10} -transformed distances were then plotted with ggplot2 as density plots, as these provided an easier visual comparison for a smaller number of regions compared to the default histograms offered by GenomicDistributions. Cell-type specificity inference and distribution across genomic classes was calculated as described in the "Functional annotation of children's H3K9me3 regions" section of Methods.

Gene set enrichment analysis of regions shared in stunting and regions shared in health was done by first assigning the regions to genes with GREAT [77] followed by passing the genes to g:Profiler [86] for identification of enriched biological terms (sources of biological terms were filtered to GO:BP, GO:MF (Gene Ontology: Molecular Function), KEGG, WikiPathways, Reactome). Significant terms were those with adjusted p-value < 0.05. Results with < 10 significantly enriched terms were plotted as bar plots showing $-\log_{10}(\text{padj})$ values. Larger sets of terms were passed as GEM files produced by g:Profiler to the EnrichmentMap app [92] in Cytoscape (3.8.0) [79], where terms were organized into networks followed by manual organization of related terms into clusters.

Integration with H3K4me3 and H3K27ac

Differential analysis results for H3K4me3 and H3K27ac data from 18-week-old and one-year-old children were obtained from Uchiyama et al. [35] and Kupkova et al. [36] respectively. These results show association of the H3K4me3 and H3K27ac peaks with a child's Δ HAZ (birth:18wk) score for 18-week-old-children and Δ HAZ (birth:1yr) score for one-year-old children. Significantly affected regions were considered those with *padj* < 0.05. Regions were extracted from the result tables and converted to BED files. Overlaps with H3K9me3 regions from this study were found with the bedtools (v2.26.0)

[61] *intersect* function with parameter *-wao* (to report original regions instead of just intersections). Proportions of overlaps were calculated as the ratio between an overlap and the size of a given region.

Gene set enrichment analysis of overlaps was done as described in the “Relationships between H3K9me3 in mothers and children” section above and histone mark/DBF enrichment analysis was performed as described in the “Functional annotation of children’s H3K9me3 regions” section above. The *userUniverse* in LOLA was set to overlaps between H3K9me3 and all upregulated H3K4me3 regions in one-year-old stunted children for H3K4me3 analysis, and to overlaps between H3K9me3 and all downregulated H3K27ac regions in one-year-old stunted children for H3K27ac analysis. For H3K27ac, only highly significant enrichments ($\text{padj} < 0.001$) were reported due to the large number of enriched histone marks and DBFs.

Additional tools used

Tidverse package (1.3.1) [93] was used to process data in R. Figures were assembled with Inkscape (1.0.2) [94], Illustrations were created with BioRender [95], and icons were obtained from the Noun Project [96].

Abbreviations

| | |
|---------------|--|
| ChIP-seq | Chromatin immunoprecipitation followed by sequencing |
| DBF | DNA binding factor |
| EED | Environmental enteric dysfunction |
| ERVs | Endogenous retroviruses |
| GOBP | Gene ontology: biological process |
| GOMF | Gene ontology: molecular function |
| H3K4me1 | Histone H3 lysine 4 monomethylation |
| H3K4me3 | Histone H3 lysine 4 trimethylation |
| H3K9me3 | Histone H3 lysine 9 trimethylation |
| H3K27ac | Histone H3 lysine 27 acetylation |
| H3K36me3 | Histone H3 lysine 36 trimethylation |
| HAZ | Height-for-age z |
| IFN- γ | Interferon gamma |
| IL8 | Interleukin 8 |
| LINE L1 | Long interspersed nuclear elements-L1 |
| NF | Normalization factor |
| PBMCs | Peripheral blood mononuclear cells |
| PC1 | First principal component |
| PC2 | Second principal component |
| PCA | Principal component analysis |
| PROVIDE | Performance of rotavirus and oral polio vaccines in developing countries (study) |
| PU.1 | Hematopoietic transcription factor PU.1 (SPI1) |
| SiH | Regions shared in health |
| SiS | Regions shared in stunting |
| SPI1 | Hematopoietic transcription factor PU.1 |
| SVA | SINE-R-VNTR-Alu |
| TEs | Transposable elements |
| TF | Transcription factor |
| TSS | Transcription start site |
| TGF β | Transforming growth factor- β |
| WAZ | Weight-for-age z |
| ZAP | Zinc-finger antiviral protein |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-023-01548-z>.

Additional file 1. Supplementary figures (S1-S14), Table S1-2.

Additional file 2: Table S3. Differential analysis results infants.

Additional file 3: Table S4. Differential analysis results mothers.

Acknowledgements

We are grateful to all the mothers and children who participated in the PROVIDE study without whom this work would not have been possible. We are also grateful to Dr. Stefan Bekiranov for suggesting analysis of transposable elements and Dr. Laura Banaszynski to providing further motivation and valuable insights regarding transposable elements. This work was funded by National Institutes of Health (Grant R01 AI043596 to W.A.P.); Biomedical Sciences Graduate Program, University of Virginia (Wagner fellowship to K.K.).

Author contributions

DA conceived of the study and supervised the analyses; SJS isolated chromatin and prepared libraries for sequencing; KK conducted the analyses; MGP-G and PAG performed Western blotting and analyzed the Western blotting results; RH and WAP, Jr, conceived of and established the PROVIDE Study. All authors contributed to writing of the manuscript. All authors read and approved the final manuscript.

Funding

Funding for this work was provided by National Institutes of Health (Grant R01 AI043596 to W.A.P) and Biomedical Sciences Graduate Program, University of Virginia (Wagner fellowship to K.K.). The funders did not have any role in design of the study and collection, analysis, and interpretation of data and in writing of the manuscript.

Availability of data and materials

All raw sequencing FASTQ files generated in this study are available in dbGaP repository, phs001073.v3.p1, <https://www.ncbi.nlm.nih.gov/gap/>. DbGaP identifiers for the samples used in this study are listed in Additional file 1: Table S1-2.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethical Review Board of icddr,b (FWA 00001468) and the Institutional Review Boards of the University of Virginia (FWA 00006183) and the University of Vermont (FWA 00000727). All samples were analyzed in a de-identified fashion. Within seven days after giving birth, screening for eligibility and study consenting occurred in the household by trained Field Research Assistants. Informed consent was obtained for all participants (trial registration: ClinicalTrials.gov NCT01375647).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, VA 22908, USA. ²Center for Public Health Genomics, University of Virginia Health System, Charlottesville, VA 22908, USA. ³Department of Biomedical Science, Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL 33431, USA. ⁴Infectious Disease Division, International Centre for Diarrhoeal Disease Research, Dhaka 1000, Bangladesh. ⁵Division of Infectious Diseases and International Health, University of Virginia Health System, Charlottesville, VA 22908, USA.

Received: 15 December 2022 Accepted: 6 August 2023

Published online: 11 August 2023

References

1. UNICEF/WHO/WBG. UNICEF/WHO/The World Bank Group joint child malnutrition estimates: levels and trends in child malnutrition: key findings of the 2021 edition [Internet]. 2021. Available from: <https://www.who.int/publications/i/item/9789240025257>
2. World Health Organization. Reducing stunting in children: equity considerations for achieving the global targets 2025. Who. 2018.
3. Bourke CD, Berkley JA, Prendergast AJ. Immune dysfunction as a cause and consequence of malnutrition. *Trends Immunol*. 2016;37(6):386–98.
4. Bourke CD, Jones KD, Prendergast AJ. Current understanding of innate immune cell dysfunction in childhood undernutrition. *Front Immunol*. 2019;10(July):1–15.
5. Alam MA, Richard SA, Fahim SM, Mahfuz M, Nahar B, Das S, et al. Impact of early-onset persistent stunting on cognitive development at 5 years of age: results from a multi-country cohort study. *PLoS ONE*. 2020;15(1):e0227839.
6. Woldehanna T, Behrman JR, Araya MW. The effect of early childhood stunting on children's cognitive achievements: Evidence from young lives Ethiopia. *Ethiop J Heal Dev = Ya'ityopya tena lemat mashet*. 2017;31(2):75.
7. Kimani-Murage EW, Kahn K, Pettifor JM, Tollman SM, Dunger DB, Gómez-Olivé XF, et al. The prevalence of stunting, overweight and obesity, and metabolic disease risk in rural South African children. *BMC Public Health*. 2010;10(1):1–13.
8. Rolfe EDL, De FGVA, Vianna CA, Gigante DP, Miranda JJ, Yudkin JS, et al. Associations of stunting in early childhood with cardiometabolic risk factors in adulthood. *PLoS ONE*. 2018;13(4):e0192196.
9. Mohammed SH, Muhammad F, Pakzad R, Alizadeh S. Socioeconomic inequality in stunting among under-5 children in Ethiopia: A decomposition analysis. *BMC Res Notes*. 2019;12(1):1–5.
10. Rizal MF, van Doorslaer E. Explaining the fall of socioeconomic inequality in childhood stunting in Indonesia. *SSM Popul Heal*. 2019;1(9):100469.
11. Bommer C, Vollmer S, Subramanian SV. How socioeconomic status moderates the stunting-age relationship in low-income and middle-income countries. *BMJ Glob Heal*. 2019;4(1):e001175.
12. Jonah CMP, Sambu WC, May JD. A comparative analysis of socioeconomic inequalities in stunting: a case of three middle-income African countries. *Arch Public Heal*. 2018;76(1):1–15.
13. Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Kung VL, Cheng J, et al. Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. *Science*. 2019;365:6449.
14. Guerrant RL, DeBoer MD, Moore SR, Scharf RJ, Lima AAM. The impoverished gut—a triple burden of diarrhoea, stunting and chronic disease. *Nat Rev Gastroenterol Hepatol*. 2013;10(4):220–9.
15. Islam MS, Zafar Ullah AN, Mainali S, Imam MA, Hasan MI. Determinants of stunting during the first 1,000 days of life in Bangladesh: a review. *Food Sci Nutr*. 2020;8(9):4685–95.
16. Mutasa K, Tome J, Rukobo S, Govha M, Mushayanembwa P, Matimba FS, et al. Stunting status and exposure to infection and inflammation in early life shape antibacterial immune cell function among Zimbabwean children. *Front Immunol*. 2022;13(13):2805.
17. Budge S, Parker AH, Hutchings PT, Garbutt C. Environmental enteric dysfunction and child stunting. *Nutr Rev*. 2019;77(4):240–53.
18. Saleh A, Syahrul S, Hadju V, Andriani I, Restika I. Role of maternal in preventing stunting: a systematic review. *Gac Sanit*. 2021;1(35):S576–82.
19. Victora CG, De Onis M, Hallal PC, Blössner M, Shrimpton R. Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics*. 2010;125(3):e473–80.
20. Martorell R, Zongrone A. Intergenerational influences on child growth and undernutrition. *Paediatr Perinat Epidemiol*. 2012;26(SUPPL. 1):302–14.
21. Dewey KG, Begum K. Long-term consequences of stunting in early life. *Matern Child Nutr*. 2011;7(SUPPL. 3):5–18.
22. WHO. Global targets 2025. Glob targets 2025. 2014.
23. Kumar M, Ji B, Babaei P, Das P, Lappa D, Ramakrishnan G, et al. Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: lessons from genome-scale metabolic modeling. *Metab Eng*. 2018;1(49):128–42.
24. Semba RD, Shardell M, Sakr Ashour FA, Moaddel R, Trehan I, Maleta KM, et al. Child stunting is associated with low circulating essential amino acids. *EBioMedicine*. 2016;1(6):246–52.
25. Moreau GB, Ramakrishnan G, Cook HL, Fox TE, Nayak U, Ma JZ, et al. Childhood growth and neurocognition are associated with distinct sets of metabolites. *EBioMedicine*. 2019;1(44):597–606.
26. Raman AS, Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Subramanian S, et al. A sparse covarying unit that describes healthy and impaired human gut microbiota development. *Science* (80-). 2019;365:6449.
27. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet*. 2012;13(2):97–109.
28. Dai Z, Ramesh V, Locasale JW. The evolving metabolic landscape of chromatin biology and epigenetics. *Nat Rev Genet*. 2020;21(12):737–53.
29. Fitz-James MH, Cavalli G. Molecular mechanisms of transgenerational epigenetic inheritance. *Nat Rev Genet*. 2022;23(6):325–41.
30. Perez MF, Lehner B. Intergenerational and transgenerational epigenetic inheritance in animals. *Nat Cell Biol*. 2019;21(2):143–51.
31. Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, et al. DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nat Commun*. 2014;5(1):1–14.
32. Peter CJ, Fischer LK, Kundakovic M, Garg P, Jakovcevski M, Dincer A, et al. DNA methylation signatures of early childhood malnutrition associated with impairments in attention and cognition. *Biol Psychiatry*. 2016;80(10):765–74.
33. Iqbal MS, Rahman S, Haque MA, Bhuyan MJ, Faruque ASG, Ahmed T. Lower intakes of protein, carbohydrate, and energy are associated with increased global DNA methylation in 2- to 3-year-old urban slum children in Bangladesh. *Matern Child Nutr*. 2019;15(3):e12815.
34. Schulze KV, Swaminathan S, Howell S, Jajoo A, Lie NC, Brown O, et al. Edematous severe acute malnutrition is characterized by hypomethylation of DNA. *Nat Commun*. 2019;10(1):1–13.
35. Uchiyama R, Kupkova K, Shetty SJ, Linford AS, Pray-Grant MG, Wagar LE, et al. Histone H3 lysine 4 methylation signature associated with human undernutrition. *Proc Natl Acad Sci USA*. 2018;115(48):E11264–73.
36. Kupkova K, Shetty SJ, Haque R, Petri WA, Auble DT. Histone H3 lysine 27 acetylation profile undergoes two global shifts in undernourished children and suggests altered one-carbon metabolism. *Clin Epigenetics*. 2021;13(1):1–15.
37. Chitrakar A, Noon M, Xiao AZ. Taming the transposon: H3K9me3 turns foe to friend in human development. *Cell Stem Cell*. 2022;29(7):1009–10.
38. Groh S, Schotta G. Silencing of endogenous retroviruses by heterochromatin. *Cell Mol Life Sci*. 2017;74(11):2055–65.
39. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* (80-). 2016;351(6277):1083–7.
40. Bakoulis S, Krautz R, Alcaraz N, Salvatore M, Andersson R. Endogenous retroviruses co-opted as divergently transcribed regulatory elements shape the regulatory landscape of embryonic stem cells. *Nucleic Acids Res*. 2022;50(4):2111–27.
41. O'Hara R, Banaszynski LA. Loss of heterochromatin at endogenous retroviruses creates competition for transcription factor binding. *bioRxiv*. 2022.
42. Hardikar AA, Satoor SN, Karandikar MS, Joglekar MV, Puranik AS, Wong W, et al. Multigenerational undernutrition increases susceptibility to obesity and diabetes that is not reversed after dietary recuperation. *Cell Metab*. 2015;22(2):312–9.
43. Klosin A, Casas E, Hidalgo-Carcedo C, Vavouri T, Lehner B. Transgenerational transmission of environmental information in *C. elegans*. *Science* (80-). 2017;356(6335):320–3.
44. Kirkpatrick BD, Colgate ER, Mychaleckyj JC, Haque R, Dickson DM, Carmolli MP, et al. The "Performance of Rotavirus and Oral Polio Vaccines in Developing Countries" (PROVIDE) study: description of methods of an interventional study designed to explore complex biologic problems. *Am J Trop Med Hyg*. 2015;92(4):744–51.
45. Howe CG, Gamble MV. Enzymatic cleavage of histone h3: A new consideration when measuring histone modifications in human samples. *Clin Epigenetics*. 2015;7(1):1–4.
46. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA*. 2010;107(50):21931–6.
47. Barral A, Pozo G, Ducrot L, Papadopoulos GL, Sauzet S, Oldfield AJ, et al. SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers. *Mol Cell*. 2022;82(4):816–832.e12.

48. Rothenberg EV, Hosokawa H, Ungerback J. Mechanisms of action of hematopoietic transcription factor PU.1 in initiation of T-cell development. *Front Immunol*. 2019;10(FEB):228.
49. Le Coz C, Nguyen DN, Su C, Nolan BE, Albrecht AV, Xhani S, et al. Constrained chromatin accessibility in PU.1-mutated agammaglobulinemia patients. *J Exp Med*. 2021;218(7):e20201750.
50. Gerber JP, Russ J, Chandrasekar V, Offermann N, Lee HM, Spear S, et al. Aberrant chromatin landscape following loss of the H3.3 chaperone Daxx in haematopoietic precursors leads to Pu.1-mediated neutrophilia and inflammation. *Nat Cell Biol*. 2021;23(12):1224–39.
51. Haque R, Mondal D, Shu J, Roy S, Kabir M, Davis AN, et al. Correlation of interferon-gamma production by peripheral blood mononuclear cells with childhood malnutrition and susceptibility to amebiasis. *Am J Trop Med Hyg*. 2007;76(2):340–4.
52. Chemudupati M, Kenney AD, Bonifati S, Zani A, McMichael TM, Wu L, et al. From APOBEC to ZAP: diverse mechanisms used by cellular restriction factors to inhibit virus infections. *Biochim Biophys Acta Mol Cell Res*. 2019;1866(3):382–94.
53. Nicetto D, Zaret KS. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. *Curr Opin Genet Dev*. 2019;1(55):1–10.
54. Notwell JH, Chung T, Heavner W, Bejerano G. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat Commun*. 2015;6(1):1–7.
55. Wagar LE, Bolen CR, Sigal N, Lopez Angel CJ, Guan L, Kirkpatrick BD, et al. Increased T cell differentiation and cytolytic function in Bangladeshi compared to American Children. *Front Immunol*. 2019;10(September):1–17.
56. Bonhoure N, Bounova G, Bernasconi D, Praz V, Lammers F, Canella D, et al. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Res*. 2014;24(7):1157–68.
57. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
58. Stolarczyk M, Reuter VP, Smith JP, Magee NE, Sheffield NC. Refgenie: a reference genome resource manager. *Gigascience*. 2020;9(2):giz149.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
60. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep*. 2019;9(1):9354.
61. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
62. Andrews S. FastQC: a quality control tool for high throughput sequence data. [Internet]. 2010. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
63. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8.
64. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92.
65. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC genome browser database: update 2006. *Nucleic Acids Res*. 2006;34(Database issue):D590.
66. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*. 2009;25(15):1952–8.
67. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1–21.
68. Wickham H. Ggplot2: elegant graphics for data analysis. 2nd ed. Cham: Springer International Publishing; 2016.
69. Garnier, Simon, Ross, Noam, Rudis, Robert, et al. viridis—Colorblind-Friendly Color Maps for R [Internet]. Available from: <https://sjmgarnier.github.io/viridis/>
70. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(W1):W187.
71. Shen L, Shao N, Liu X, Nestler E. Ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics*. 2014;15(1):1–14.
72. Kupkova K, Mosquera JV, Smith JP, Stolarczyk M, Danehy TL, Lawson JT, et al. GenomicDistributions: fast analysis of genomic intervals with Bioconductor. *BMC Genomics*. 2022;23(1):1–6.
73. Gao T, Qian J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res*. 2020;48(D1):D58–64.
74. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*. 2016;32(4):587–9.
75. Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, et al. Cistrome data browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res*. 2017;45(D1):D658–62.
76. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, et al. Cistrome data browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res*. 2019;47(D1):D729–35.
77. McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495–501.
78. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. 2011;6(7):e21800.
79. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
80. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46(D1):D794–801.
81. Bolstad B. preprocessCore: A collection of pre-processing functions [Internet]. 2021. Available from: <https://github.com/bmbolstad/preprocessCore>
82. Rainer J. EnsDb.Hsapiens.v75: Ensembl based annotation package. 2017.
83. Rainer J, Gatto L, Weichenberger CX. ensemblDb: an R package to create and use ensembl-based annotation resources. *Bioinformatics*. 2019;35(17):3151–3.
84. Morgan M, Rainer J. AnnotationFilter: Facilities for Filtering Bioconductor Annotation Resources [Internet]. 2020. Available from: <https://github.com/Bioconductor/AnnotationFilter>
85. Arora S, Morgan M, Carlson M, Pagès H. GenomInfoDb: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style [Internet]. 2021. Available from: <https://bioconductor.org/packages/GenomInfoDb>
86. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1):W191–8.
87. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. 2019;14(2):482–517.
88. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015 [Internet]. Available from: <http://www.repeatmasker.org>
89. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017;33(18):2938–40.
90. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13.
91. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847–9.
92. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. Ravasi T, editor. *PLoS ONE*. 2010;5(11):e13984.
93. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4(43):1686.
94. Draw Freely | Inkscape [Internet]. [cited 2022 Dec 14]. Available from: <https://inkscape.org/>

95. BioRender [Internet]. [cited 2021 Feb 18]. Available from: <https://biorender.com/>
96. Noun Project: Free Icons & Stock Photos for Everything [Internet]. [cited 2022 Dec 14]. Available from: <https://thenounproject.com/>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

